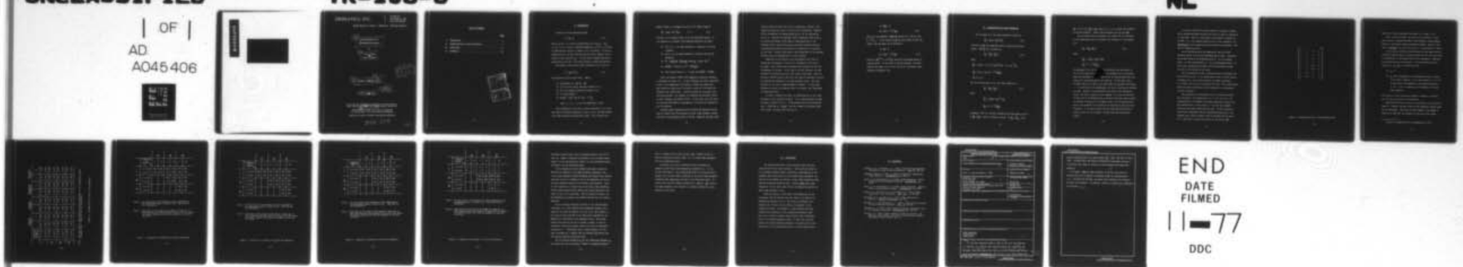
AD-A045 406 DESMATICS INC STATE COLLEGE PA F/G 12/1
RIDGE REGRESSION FOR NONSTANDARDIZED MODELS.(U)
OCT 77 T L KING, D E SMITH N00014-75-C-1054
UNCLASSIFIED TR-106-6 NL

1 OF 1
AD
A045406



END
DATE
FILMED
11-77
DDC

DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

6

RIDGE REGRESSION FOR NONSTANDARDIZED MODELS.

by

10

Terry L./King
[redacted]
Dennis E./Smith

9 TECHNICAL REP[redacted]T, NO. 14 TR- 1Ø6-6

12 23p.

11 Oct[redacted] 1977

This study was supported by the Office of Naval Research
under Contract No. NØØØ14-75-C-1Ø54, Task No. NR 042-334
15

1473
391 156

LB

TABLE OF CONTENTS

I. INTRODUCTION

Consider the usual regression model

$$\underline{y} = \underline{X\beta} + \underline{\varepsilon} \tag{1.1}$$

where $\underline{y}$ is an n x 1 vector of observations, $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ is a (p + 1) x 1 vector of unknown parameters, $\underline{\varepsilon}$ is an n x 1 vector of random errors, and $\underline{X}$ is a fixed n x (p + 1) matrix. It will be assumed that $\underline{X}$ is of full rank and that the first column of $\underline{X}$ is a column of ones, denoted by $\underline{1}$. It will also be assumed that $\underline{\varepsilon}$ has expectation $\underline{0}$, Var $(\underline{\varepsilon}) = \sigma^2 \underline{I}$, and $\underline{\varepsilon}$ follows a normal distribution.

The ordinary least squares (OLS) estimator of $\underline{\beta}$ is given by

$$\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y} \tag{1.2}$$

The properties of $\hat{\underline{\beta}}$ are well known. Namely,

(1) $\hat{\underline{\beta}}$ minimizes $(\underline{y} - \underline{X\beta})'(\underline{y} - \underline{X\beta})$

(2) $\hat{\underline{\beta}}$ is the best linear unbiased estimator of $\underline{\beta}$

(3) $\hat{\underline{\beta}}$ is the maximum likelihood estimator of $\underline{\beta}$

(4) $\hat{\underline{\beta}} \sim N(\underline{\beta}, \sigma^2(\underline{X}'\underline{X})^{-1})$

(5) $\mathrm{MSE}(\hat{\underline{\beta}}) = E[(\hat{\underline{\beta}} - \underline{\beta})'(\hat{\underline{\beta}} - \underline{\beta})] = \sigma^2 \Sigma \frac{1}{\lambda_i}$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\underline{X}'\underline{X}$.

Upon examination of the last of these properties, it is clear that when the minimum eigenvalue is close to zero, the mean squared error (MSE) becomes unsatisfactorily large. This led Hoerl and

Kennard (1970a) to recommend the use of the "ridge estimator"

$$\underline{\hat{\beta}}^* = (\underline{X}'\underline{X} + k\underline{I})^{-1}\underline{X}'\underline{y}, \qquad k > 0 \tag{1.3}$$

when $\underline{X}'\underline{X}$, in correlation form, is an ill-conditioned matrix. If k is treated as a constant, the following properties are known:

(1) For $k = 0$, the ridge estimator is identical to the OLS estimator.

(2) For $k > 0$, the ridge estimator is shorter than the OLS estimator.

(3) $\underline{\hat{\beta}}^* \sim N(\underline{W}_k\underline{X}'\underline{X\beta}, \sigma^2\underline{W}_k\underline{X}'\underline{XW}_k)$ where $\underline{W}_k = (\underline{X}'\underline{X} + k\underline{I})^{-1}$.

(4) $MSE(\underline{\hat{\beta}}^*) = \sigma^2\Sigma\lambda_i/(\lambda_i + k)^2 + k^2\underline{\beta}'\underline{W}_k{}^2\underline{\beta}$.

(5) There always exists a $k > 0$ such that $MSE(\underline{\hat{\beta}}^*) < MSE(\underline{\hat{\beta}})$.

Hoerl and Kennard (1970b) also suggested a graphical technique to determine the value of k. In this technique, one first constructs a plot of the components of $\underline{\hat{\beta}}^*$ versus k, called the ridge trace. Then using this ridge trace, he chooses a value of k for which the estimates have "stabilized". Analytic methods for choosing k have also been proposed. For example, see McDonald and Galarneau (1975), or Hoerl, Kennard, and Baldwin (1975). In any event, once the value of k has been determined in some manner, (1.3) gives the estimate of $\underline{\beta}$ to be reported.

Research papers appearing since the Hoerl and Kennard articles have not always used this estimator as their ridge estimator because the form of the regression model is varied. Marquardt and Snee (1975)

advocate using the model with only $\underline{X}$ standardized. However, they suggest choosing the value of k with $\underline{y}$ also standardized. Hemmerle (1975) standardizes the design matrix $\underline{X}$, but not the observation vector, $\underline{y}$. McDonald and Galarneau (1975) standardize both $\underline{X}$ and $\underline{y}$, but Guilkey and Murphy (1975) make no standardizations at all. Obenchain (1975) centers both $\underline{X}$ and $\underline{y}$ and warns against scaling or reparameterizing unless the results are ultimately to be reported in that form. Thisted (1976) gives a good discussion of the problems involved with the standardization of the variables.

Comparison of the results of one investigator with those of another may be hampered by the lack of uniformity in the form of the model. Most authors have recognized this problem, but have not investigated it further. In the case of the OLS solution, the same estimator is produced from any of the forms of the model. Thus, by solving a problem stated in any form, the same OLS estimator of $\underline{\beta}$ will be produced when measured in the same parameter space. This nice property is not true of ridge estimators, however. In this paper, formulae are given for different forms of the model, and comparisons are made among them.

In order to examine the effect of standardization on the ridge estimator, first consider the model. If no transformations are made, the model is given in (1.1). If the $\underline{X}$ matrix and $\underline{\beta}$ are partitioned as $\underline{X} = [\underline{1} \vdots \underline{G}]$ and $\underline{\beta} = (\beta_0, \underline{\beta}_G')'$, and the columns of $\underline{G}$ centered about their means, the model can be written as

$$\underline{y} = \underline{C}\underline{G}\underline{\beta}_G + \underline{\varepsilon}$$

$$\beta_0 = E[\bar{y}] - n^{-1}\underline{1}'\underline{G}\underline{\beta}_G \tag{1.4}$$

where $\underline{C}$ is the symmetric idempotent matrix $(\underline{I} - n^{-1}\underline{1}\,\underline{1}')$, and $\bar{y} = n^{-1}\Sigma y_i$. If the vectors in $\underline{CG}$ are also scaled to have unit length, then the model can be written as

$$\underline{y} = \underline{H}\underline{\gamma} + \underline{\varepsilon}$$

$$\beta_0 = E[\bar{y}] - n^{-1}\underline{1}'\underline{G}\underline{\beta}_G \tag{1.5}$$

where $\underline{H} = \underline{CGD}^{-1/2}$, $\underline{\gamma} = \underline{D}^{1/2}\underline{\beta}_G$, and $\underline{D}$ is the diagonal matrix of scaling factors. In the case of the OLS estimator, the three forms of the model, (1.1), (1.4), and (1.5), are known to give equivalent estimators of $\underline{\beta}$.

II. STANDARDIZATION IN RIDGE ESTIMATION

For the model (1.1) the ridge estimator is given by

$$\underline{\hat{\beta}}_1^* = (\underline{X}'\underline{X} + k_1\underline{I})^{-1}\underline{X}'\underline{y} \tag{2.1}$$

In order to make the comparison easier, partition $\underline{X}$ and $\underline{\beta}$ as before. Then $\underline{\hat{\beta}}_1^*$ can be written as

$$\underline{\hat{\beta}}_1^* = (\hat{\beta}_{01}^*, \underline{\hat{\beta}}_{G1}^{*\prime})' \tag{2.2}$$

where

$$\underline{\hat{\beta}}_{G1}^* = (\underline{G}'(\underline{I} - (n + k_1)^{-1}\underline{J})\underline{G})^{-1}\underline{G}'(\underline{I} - (n + k_1)^{-1}\underline{J})\underline{y}$$

$$\hat{\beta}_{01}^* = (n/(n + k_1))(\bar{y} - n^{-1}\underline{1}'\underline{G\hat{\beta}}_{G1}^*)$$

and $\underline{J} = \underline{1}\ \underline{1}'$.

For the centered model (1.4), the ridge estimator is

$$\underline{\hat{\beta}}_2^* = (\hat{\beta}_{02}^*, \underline{\hat{\beta}}_{G2}^{*\prime})' \tag{2.3}$$

where

$$\underline{\hat{\beta}}_{G2}^* = (\underline{G}'\underline{CG} + k_2\underline{I})^{-1}\underline{G}'\underline{Cy}$$

$$\hat{\beta}_{02}^* = \bar{y} - n^{-1}\underline{1}'\underline{G\hat{\beta}}_{G2}^*$$

In general, this is a distinct estimator from that given in (2.2). If $\underline{\hat{\beta}}_{G1}^* \neq \underline{\hat{\beta}}_{G2}^*$, there is nothing to prove. If $\underline{\hat{\beta}}_{G1}^* = \underline{\hat{\beta}}_{G2}^*$, then

$\hat{\beta}^*_{01} = (n/(n + k_1))\hat{\beta}^*_{02} \neq \hat{\beta}^*_{02}$ unless $k_1 = 0$; i.e., unless the estimator is the OLS estimator. Hence, these estimators are not the same.

For the third form of the model (1.5), which will be referred to as the standardized form of the model, the estimator of the original $\underline{\beta}$ is

$$\hat{\underline{\beta}}^*_3 = (\hat{\beta}^*_{03}, \hat{\underline{\beta}}^*_{G3}{}')' \tag{2.4}$$

where

$$\hat{\underline{\beta}}^*_{G3} = (\underline{G}'\underline{C}\underline{G} + k_3\underline{D})^{-1}\underline{G}'\underline{C}\underline{y}$$

$$\hat{\beta}^*_{03} = \bar{y} - n^{-1}\underline{1}'\underline{G}\hat{\underline{\beta}}_{G3}$$

The estimator (2.4) can be seen to be different from that given in (2.2) in the same way as [illegible] was. The estimators for the centered model are equivalent t[illegible]imators for the standardized model only when $\underline{D}$ is a scalar multiple of the identity. In this case, there do exist values of k_2 and k_3 that would make the estimators equivalent.

To this point in the development, the error structure has remained the same. However, if transformations are made on the observation vector, the error structure is also changed. It is not too difficult to show that if the design matrix is centered, the same ridge estimator is produced by using $\underline{y}$ in its original form or in its centered form, even if the equations to be solved do not take into account the error structure. It can also be shown that the centering and scaling of $\underline{y}$ has no effect on the estimator if $\underline{X}$ has also been centered and scaled.

In order to evaluate the three classes of estimators, squared error (squared Euclidean distance) was used as a basis of comparison in this study. While it is possible to find explicit formulae for the MSE (i.e., the expected squared error) of each estimator, the distribution of the squared error also should be investigated. This led to a simulation study.

As the discussion above has emphasized, there are three parameter spaces in which such evaluation may be made. Comparisons can be made based on the parameterization of (1) the original model, (2) the centered model, or (3) the standardized model. In this study, comparisons were made in all three parameter spaces and also in the observation ($\underline{y}$) space.

For this simulation study, a design discussed in Marquardt and Snee (1975) was used. In the design matrix there are three vectors orthogonal to $\underline{1}$. Two of these vectors are highly correlated, ($r = 0.989$), and the other correlations are zero. This design matrix, which has eight observations and four parameters to be estimated, is given in Figure 1.

The eigenvalues and eigenvectors for the correlation matrix were calculated. The eigenvectors $(-1, -1, 0)'$ and $(-1, 1, 0)'$ corresponding to the maximum and minimum eigenvalues respectively, were multiplied by σ and 3σ, where $\sigma = 0.8$, the same standard deviation value used by Marquardt and Snee. The four resulting vectors were transformed from the standardized space back to the original space, and the constant term in the model was set equal to 0. Using each of these four vectors as the true $\underline{\beta}$, 1000

$$
\begin{bmatrix}
1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 \\
1 & 1 & 1 & 1 \\
1 & -1 & -0.8 & -1 \\
1 & 1 & 0.8 & -1 \\
1 & 0.8 & 1 & 1 \\
1 & -0.8 & -1 & 1
\end{bmatrix}
$$

Figure 1: Design Matrix Used in the Simulation Study

replicates of $\underline{y}$ were generated from model (1.1) using $\sigma = 0.8$.

Since σ and $\underline{\beta}$ are known, the value of k was chosen by a computer routine to minimize the MSE of each estimator, using formulae similar to that given in Hoerl and Kennard (1970a). Hence, in each case, the optimal[1] value of k was used, i.e., conditional on σ and $\underline{\beta}$.

At this point, the MSE of each estimator in each space could be calculated theoretically, but doing so would not tell very much about the distribution of the squared error of the estimators. Therefore, the simulation study was conducted. In this simulation the true value of $\underline{\beta} = (\beta_0, \underline{\beta}_G')'$ is defined such that

(1) $\beta_0 = 0$

(2) $\underline{\beta}_G$, when transformed to the standardized space, is equal to $\alpha\underline{v}$, where α denotes σ or 3σ, and $\underline{v}$ denotes $(-1, -1, 0)'$, an eigenvector corresponding to the maximum eigenvalue, or $(-1, 1, 0)'$, an eigenvector corresponding to the minimum eigenvalue.

The values of α and $\underline{v}$ are used in Figures 2 through 6 to indicate the true value of $\underline{\beta}$ used in the simulation.

Estimated MSE values for each of the estimators are presented in Figure 2. Figures 3 through 6 show how the estimators compare based on the simulations. Each entry in these figures is the number of times out of 1000 that the estimator at the top of that column

---

[1] Subject to computing error not exceeding 1.0×10^{-5}.

	Multiple α / Vector $\underline{v}$	Original Parameters σ	Original Parameters 3σ	Centered Parameters σ	Centered Parameters 3σ	Standardized Parameters σ	Standardized Parameters 3σ	Observation Vector σ	Observation Vector 3σ
$\hat{\underline{\beta}}$	(-1,-1,0)'	12.997	12.997	12.866	12.866	93.758	93.758	4.102	4.102
	(-1,1,0)'	12.997	12.997	12.866	12.866	93.758	93.758	4.102	4.102
$\hat{\underline{\beta}}^*_1$	(-1,-1,0)'	0.227	0.374	0.160	0.273	1.210	2.057	5.856	5.835
	(-1,1,0)'	1.280	6.492	1.280	6.362	9.314	46.409	7.958	4.437
$\hat{\underline{\beta}}^*_2$	(-1,-1,0)'	0.295	0.406	0.164	0.275	1.245	2.075	5.531	5.739
	(-1,1,0)'	1.411	6.494	1.280	6.363	9.316	46.415	6.887	4.437
$\hat{\underline{\beta}}^*_3$	(-1,-1,0)'	0.291	0.404	0.159	0.272	1.211	2.054	5.557	5.754
	(-1,1,0)'	1.411	6.494	1.280	6.363	9.315	46.412	6.886	4.437

Note: The true value of the parameter $\underline{\beta}$, when transformed to the standardized space, is equal to $\alpha\underline{v}$. (See text for full discussion.)

Figure 2: Estimated MSE for Each of the Estimators

		$\hat{\underline{\beta}}^*_1$		$\hat{\underline{\beta}}^*_2$		$\hat{\underline{\beta}}^*_3$	
	Multiple α / Vector $\underline{v}$	σ	3σ	σ	3σ	σ	3σ
$\hat{\underline{\beta}}$	(-1,-1,0)'	981	965	979	954	979	954
	(-1,1,0)'	787	573	771	573	771	573
$\hat{\underline{\beta}}^*_1$	(-1,-1,0)'			178	150	201	195
	(-1,1,0)'			49	296	55	305
$\hat{\underline{\beta}}^*_2$	(-1,-1,0)'					748	786
	(-1,1,0)'					606	597

Note 1: The true value of the parameter $\underline{\beta}$, when transformed to the standardized space, is equal to $\alpha\underline{v}$. (See text for full discussion.)

Note 2: Each entry in this figure is the number of times out of 1000 replicates that the estimator at the top of the column had smaller squared error than the estimator at the left of the row.

Figure 3: Comparison of Estimates of $\underline{\beta}$ from the Simulation.

	Multiple α / vector $\underline{v}$	$\hat{\underline{\beta}}_1^*$ σ	$\hat{\underline{\beta}}_1^*$ 3σ	$\hat{\underline{\beta}}_2^*$ σ	$\hat{\underline{\beta}}_2^*$ 3σ	$\hat{\underline{\beta}}_3^*$ σ	$\hat{\underline{\beta}}_3^*$ 3σ
$\hat{\underline{\beta}}$	(-1,-1,0)'	969	953	979	954	979	954
	(-1,1,0)'	771	573	771	573	771	573
$\hat{\underline{\beta}}_1^*$	(-1,-1,0)'			508	470	591	605
	(-1,1,0)'			459	575	477	619
$\hat{\underline{\beta}}_2^*$	(-1,-1,0)'					748	786
	(-1,1,0)'					609	597

Note 1: The true value of the parameter $\underline{\beta}$, when transformed to the standardized space, is equal to $\alpha\underline{v}$. (See text for full discussion.)

Note 2: Each entry in this figure is the number of times out of 1000 replicates that the estimator at the top of the column had smaller squared error than the estimator at the left of the row.

Figure 4: Comparison of Estimates of $\underline{\beta}_G$ from the Simulation.

		$\hat{\underline{\beta}}^*_1$		$\hat{\underline{\beta}}^*_2$		$\hat{\underline{\beta}}^*_3$	
	Multiple α / Vector $\underline{v}$	σ	3σ	σ	3σ	σ	3σ
$\hat{\underline{\beta}}$	(-1,-1,0)'	971	953	979	954	980	954
	(-1,1,0)'	771	573	771	573	771	573
$\hat{\underline{\beta}}^*_1$	(-1,-1,0)'			495	467	580	610
	(-1,1,0)'			452	575	472	622
$\hat{\underline{\beta}}^*_2$	(-1,-1,0)'					755	791
	(-1,1,0)'					609	612

Note 1: *The true value of the parameter* $\underline{\beta}$, when transformed to the standardized space, is equal to $\alpha\underline{v}$. (See text for full discussion.)

Note 2: Each entry in this figure is the number of times out of 1000 replicates that the estimator at the top of the column had smaller squared error than the estimator at the left of the row.

Figure 5: Comparison of Estimates of $\underline{\gamma}$ from the Simulation.

		$\hat{\underline{\beta}}_1^*$		$\hat{\underline{\beta}}_2^*$		$\hat{\underline{\beta}}_3^*$	
	Multiple α / Vector $\underline{v}$	σ	3σ	σ	3σ	σ	3σ
$\hat{\underline{\beta}}$	(-1,-1,0)'	0	0	0	0	0	0
	(-1,1,0)'	0	0	0	0	0	0
$\hat{\underline{\beta}}_1^*$	(-1,-1,0)'			1000	1000	1000	1000
	(-1,1,0)'			1000	1000	1000	949
$\hat{\underline{\beta}}_2^*$	(-1,-1,0)'					0	0
	(-1,1,0)'					545	0

Note 1: The true value of the parameter $\underline{\beta}$, when transformed to the standardized space, is equal to $\alpha\underline{v}$. (See text for full discussion.)

Note 2: Each entry in this figure is the number of times out of 1000 replicates that the estimator at the top of the column had smaller squared error than the estimator at the left of the row.

Figure 6: Comparison of Estimates of $\underline{y}$ from the Simulation.

had smaller squared error than the estimator given at the left of that row. Figure 3 compares the estimators in the original space, Figure 4 in the centered space, Figure 5 in the standardized space, and Figure 6 in the observation space.

In the simulation the OLS estimator performed as expected. When OLS was compared to the ridge estimators separately, each of the ridge estimators yielded estimates with squared error smaller than that of OLS in a minimum of 57% of the simulations in each of the three parameter spaces for all values of $\underline{\beta}$ considered. Since the OLS estimator is a special case of all three ridge estimators, and the ridge estimators were chosen with optimal properties, this result is not at all surprising. Also as expected in theory, the OLS estimates of $\underline{y}$ produced the smallest squared error for every $\underline{y}$ generated.

For the original parameter $\underline{\beta}$ (Figure 3) the ridge estimates from model (1.1), when compared with transformed estimates from either of the other two models, (1.4) and (1.5), were closest to the true $\underline{\beta}$ in more than 69% of the simulations regardless of the magnitude or orientation of the parameter vector. Since these results were based on the use of optimal k values, it does not necessarily follow that similar results will hold for nonoptimal (stochastic) k . Nonetheless, when a ridge estimator is to be used to estimate $\underline{\beta}$, it appears that the estimator $\hat{\underline{\beta}}_1^*$ derived from the original model may perform the best.

For the centered parameter $\underline{\beta}_G$ and the standardized parameter $\underline{\gamma}$, the results were not as striking. However, by examining Figures 4

and 5 it appears that in each of these cases it might be well to base the estimation process on $\hat{\underline{\beta}}_3^*$, i.e., by using ridge estimation with the standardized model.

An estimate of $\underline{y}$ can be obtained by first estimating the parameter $\underline{\beta}$ and then pre-multiplying the estimate by $\underline{X}$. As can be seen from Figure 6, the estimator $\hat{\underline{\beta}}_2^*$ based on the centered model appears to be the most likely candidate of the three ridge estimators considered when comparisons are made in the observation space. Of course, since the OLS estimator minimizes $(\underline{y} - \underline{X\beta})'(\underline{y} - \underline{X\beta})$, none of the ridge estimators ever resulted in a smaller squared error when measured on this basis.

III. CONCLUSIONS

The research described in this technical report has been concerned with three ridge estimators. Since each was chosen to be an optimal estimator within a particular transformation of the parameter space with respect to the problem being considered, the results do not address the question of when ridge estimation rather than OLS estimation should be used. Instead, given that ridge regression is to be used, these results do address the question of which ridge estimator to use.

While the results of this small scale simulation are not conclusive, they do indicate that care needs to be taken in deciding which estimator to use. The choice of estimator will depend on the criteria used to measure the goodness of the estimation process. For example, a ridge estimator with good squared error properties in the standardized parameter space may not do as well as another ridge estimator when transformed back to the original parameter space. In this case, the data analyst must decide whether measurement of squared error is more appropriate in the standardized space or in the original space.

IV. REFERENCES

Guilkey, D. K., and Murphy, J. L. (1975) Directed ridge regression techniques in cases of multicollinearity. JASA, 70, 769-776.

Hemmerle, William J. (1975) An explicit solution for generalized ridge regression. Technometrics, 17, 309-314.

Hoerl, A. E. and Kennard, R. W. (1970a) Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1970b) Ridge regression: applications to nonorthogonal problems. Technometrics, 12, 69-82.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975) Ridge regression: some simulations. Comm. Stat., 4, 105-123.

Marquardt, D. W. and Snee, R. D. (1975) Ridge regression in practice. American Statistician, 29, 3-20.

McDonald, G. C. and Galarneau, D. I. (1975) A Monte-Carlo evaluation of some ridge-type estimators. JASA, 70, 407-416.

Obenchain, R. L. (1975) Ridge analysis following a preliminary test of a shrunken hypothesis. Technometrics, 17, 431-441.

Thisted, R. A. (1976) Ridge regression, minimax estimation, and empirical Bayes methods. Division of Biostatistics, Stanford University, Technical Report No. 28.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 106-6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) RIDGE REGRESSION FOR NONSTANDARDIZED MODELS		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Terry L. King and Dennis E. Smith		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-1054
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P. O. Box 618 State College, Pa. 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-334
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics and Probability Program (Code 436) Office of Naval Research Arlington, VA 22217		12. REPORT DATE October 1977 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ridge regression
Standardization
Linear model

inverse

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

epsilon

In the usual regression model $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ with $\underline{X}$ of full rank and $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2\underline{I})$, the ordinary least squares estimator $\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$ has

sigma squared   beta

approx

extremely large mean square error when $\underline{X}$ is an "ill-conditioned" matrix.   next page

Hoerl and Kennard [Technometrics, Vol. 12, No. 1, pp. 55-67 (1970)] have

DD FORM 1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

proposed replacing $\hat{\underline{\beta}}$ by the ridge estimator $\hat{\underline{\beta}}_k^* = (\underline{X}'\underline{X} + k\underline{I})^{-1}\underline{X}'\underline{y}$ in these cases. Although Hoerl and Kennard considered the regression problem in correlation form, there is nothing in their procedure that makes this mandatory.

This paper compares ridge estimators for $\underline{\beta}$ beta that arise when the biasing factor (k) is applied at different stages of standardization (i.e., centering and scaling), and shows which estimators are identical and which are different. In addition, results of a small-scale simulation are discussed.